Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces

Aapo Hyvärinen and Patrik Hoyer
Helsinki University of Technology
Laboratory of Computer and Information Science
P.O. Box 5400, FIN-02015 HUT, Finland
aapo.hyvarinen@hut.fi, patrik.hoyer@hut.fi

To appear in Neural Computation

August 16, 1999

Abstract

Olshausen and Field (1996) applied the principle of independence maximization by sparse coding to extract features from natural images. This leads to the emergence of oriented linear filters that have simultaneous localization in space and in frequency, thus resembling Gabor functions and simple cell receptive fields. In this paper, we show that the same principle of independence maximization can explain the emergence of phase and shift invariant features, similar to those found in complex cells. This new kind of emergence is obtained by maximizing the independence between norms of projections on linear subspaces (instead of the independence of simple linear filter outputs). The norms of the projections on such 'independent feature subspaces' then indicate the values of invariant features.

1 Introduction

A fundamental approach in signal processing is to design a statistical generative model of the observed signals. Such an approach is also useful for modeling the properties of neurons in primary sensory areas. Modeling visual data by a simple linear generative model, Olshausen and Field (1996) showed that the principle of maximizing the sparseness (or nongaussianity) of the underlying image components is enough to explain the emergence of Gabor-like filters that resemble the receptive fields of simple cells in mammalian primary visual cortex (V1). Maximizing sparseness is in this context equivalent to maximizing the independence of the image components (Comon, 1994; Bell and Sejnowski, 1997; Olshausen and Field, 1996). We show in this paper that this same principle can also explain the emergence of phase and shift invariant features, i.e. the principal properties of complex cells in V1. Using the method of feature subspaces (Kohonen, 1995; Kohonen, 1996), we model the response of a complex cell as the norm of the projection of the input vector (image patch) onto a linear subspace, which is equivalent to the classical energy models. Then we maximize the independence between the norms of such projections, or energies. Thus we obtain features that are localized in space, oriented, and bandpass (selective to scale/frequency), like those given by simple cells, or Gabor analysis. In contrast to simple linear filters, however, the obtained features also show emergence of phase invariance and (limited) shift or translation invariance. Phase invariance means that the response does not depend on the Fourier-phase of the stimulus: the response is the same for a white bar and a black bar, as well as for a bar and an edge. Limited shift invariance means that a near-maximum response can be elicited by identical bars or edges at slightly different locations. These two latter properties closely parallel the properties that distinguish complex cells from simple cells in V1. Maximizing the independence, or equivalently, the sparseness of the norms of the projections to feature subspaces thus allows for the emergence of exactly those invariances that are encountered in complex cells, indicating their fundamental importance in image data.

2 Independent component analysis of image data

The basic models that we consider here express a static monochrome image $I(x, y)$ as a linear superposition of some features or basis functions $b_i(x, y)$:

$$I(x,y) = \sum_{i=1}^{m} b_i(x,y) s_i \tag{1}$$

where the s_i are stochastic coefficients, different for each image $I(x, y)$. The crucial assumption here is that the s_i are nongaussian, and mutually independent. This type of decomposition is called independent component analysis (ICA) (Comon, 1994; Bell and Sejnowski, 1997; Hyvärinen and Oja, 1997), or from an alternative viewpoint, sparse coding (Olshausen and Field, 1996; Olshausen and Field, 1997).

Estimation of the model in Eq. (1) consists of determining the values of s_i and $b_i(x, y)$ for all i and (x, y), given a sufficient number of observations of images, or in practice, image patches $I(x, y)$. We restrict ourselves here to the basic case where the $b_i(x, y)$ form an invertible linear system. Then we can invert the system as

$$s_i = < w_i, I > \tag{2}$$

where the w_i denote the inverse filters, and $< w_i, I >= \sum_{x,y} w_i(x, y)I(x, y)$ denotes the dot-product. The $w_i(x, y)$ can then be identified as the receptive fields of the model simple cells, and the s_i are their activities when presented with a given image patch $I(x, y)$. Olshausen and Field (1996) showed that when this model is estimated with input data consisting of patches of natural scenes, the obtained filters $w_i(x, y)$ have the three principal properties of simple cells in V1: they are localized, oriented, and bandpass. Van Hateren and van der Schaaf (1998) compared quantitatively the obtained filters $w_i(x, y)$ with those measured by single-cell recordings of the macaque cortex, and found a good match for most of the parameters.

3 Decomposition into independent feature subspaces

3.1 Introduction

In addition to the essentially linear simple cells, another important class of cells in V1 is complex cells. Complex cells share the above-mentioned properties of simple cells but have the two principal distinguishing properties of phase invariance and (limited) shift invariance (Hubel and Wiesel, 1962; Pollen and Ronner, 1983), at least for the preferred orientation and frequency. Note that although invariance with respect to shift and global Fourier phase are equivalent, they are different properties when phase is computed from a local Fourier transform. Another distinguishing property of complex cells is that the receptive fields are larger than in simple cells, but this difference is only quantitative, and of less consequence here. For more details see e.g. (Heeger, 1992; Pollen and Ronner, 1983; Mel et al., 1998). To this date, very few attempts have been made to formulate a statistical model that would explain the emergence of the properties of visual complex cells. It is simple to see why ICA as in Eq. (1) cannot be directly used for modeling complex cells. This is due to the fact that in that model the activations of the neurons s_i can be used to linearly reconstruct the image $I(x, y)$, which is not true for complex cells due to their two principal properties of phase invariance and shift invariance: The responses of complex cells do not give the phase or the exact position of the stimulus, at least not as a linear function as in Eq. (1). (See (von der Malsburg et al., 1998) for a nonlinear reconstruction of the image from complex cell responses.)

The purpose of this paper is to explain the emergence of phase and shift invariant features using a modification of the ICA model. The modification is based on combining the technique of multidimensional independent component analysis (Cardoso, 1998) and the principle of invariant-feature subspaces (Kohonen, 1995; Kohonen, 1996). We first describe these two recently developed techniques.

3.2 Invariant feature subspaces

The classical approach for feature extraction is to use linear transformations, or filters. The presence of a given feature is detected by computing the dot-product of input data with a given feature vector. For example, wavelet, Gabor, and Fourier transforms, as well as most models of V1 simple cells, use such linear features. The problem with linear features is, however, that they necessarily lack any invariance with respect to such transformations as spatial shift or change in (local) Fourier phase (Pollen and Ronner, 1983; Kohonen, 1996).

Kohonen (1996) developed the principle of invariant-feature subspaces as an abstract approach to representing features with some invariances. The principle of invariant-feature subspaces states that one may consider an invariant feature as a linear subspace in a feature space. The value of the invariant, higher-order feature is given by (the square of) the norm of the projection of the given data point on that subspace, which is typically spanned by lower-order features.

A feature subspace, as any linear subspace, can always be represented by a set of orthogonal basis vectors, say $w_i(x, y), i = 1, ..., n$, where n is the dimension of the subspace. Then the value $F(I)$ of the feature F with input vector $I(x, y)$ is given by

$$F(I) = \sum_{i=1}^{n} < w_i, I >^2 \tag{3}$$

(For simplicity of notation and terminology, we do not distinguish clearly the norm and the square of the norm in this paper). In fact, this is equivalent to computing the distance between the input vector $I(x, y)$ and a general linear combination of the basis vectors (filters) $w_i(x, y)$ of the feature subspace (Kohonen, 1996). A graphical depiction of feature subspaces is given in Fig. 1.

In (Kohonen, 1996), it was shown that this principle, when combined with competitive learning techniques, can lead to emergence of invariant image features.

3.3 Multidimensional independent component analysis

In multidimensional independent component analysis (Cardoso, 1998), a linear generative model as in Eq. (1) is assumed. In contrast to ordinary ICA, however, the components (responses) s_i are not assumed to be all mutually independent. Instead, it is assumed that the s_i can be divided into couples, triplets or in general n-tuples, such that the s_i inside a given n-tuple may be dependent on each other, but dependencies between different n-tuples are not allowed.

Every n-tuple of s_i corresponds to n basis vectors $b_i(x, y)$. We call a subspace spanned by a set of n such basis vectors an 'independent (feature) subspace'. In general, the dimensionality of each independent subspace need not be equal, but we assume so for simplicity.

The model can be simplified by two additional assumptions. First, even though the components s_i are not all independent, we can always define them so that they are uncorrelated, and of unit variance. In fact, linear dependencies inside a given n-tuple of dependent components could always be removed by a linear transformation. Second, we can assume that the data is whitened (sphered); this can be always accomplished by e.g. PCA (Comon, 1994). Whitening is a conventional preprocessing step in ordinary ICA, where it makes the basis vectors b_i orthogonal (Comon, 1994; Hyvärinen and Oja, 1997), if we ignore any finite-sample effects.

These two assumptions imply that the b_i are orthonormal, and that we can take $b_i = w_i$ as in ordinary ICA with whitened data. In particular, the independent subspaces become orthogonal after whitening. These facts follow directly from the proof in (Comon, 1994), which applies here as well, due to our above assumptions.

Let us denote by J the number of independent feature subspaces, and by $S_j, j = 1, ..., J$ the set of the indices of the s_i belonging to the subspace of index j. Assume that the data consists of K observed image patches $I_k(x, y), k = 1, ..., K$. Then we can express the likelihood L of the data given the model as follows

$$\begin{aligned} & L(I_k(x,y), k = 1...K; w_i(x,y), i = 1...m) \\ &= \prod_{k=1}^{K} [|\det \mathbf{W}| \prod_{j=1}^{J} p_j(< w_i, I_k >, i \in S_j)] \end{aligned} \tag{4}$$

where $p_j(.)$, which is a function of the n arguments $< w_i, I_k >, i \in S_j$, gives the probability density inside the j-th n-tuple of s_i, and $\mathbf{W}$ is a matrix containing the filters $w_i(x, y)$ as its columns. The term $|\det \mathbf{W}|$ appears here as in any expression of the probability density of a transformation, giving the change in volume produced by the linear transformation, see e.g. (Pham et al., 1992).

The n-dimensional probability density $p_j(.)$ is not specified in advance in the general definition of multidimensional ICA (Cardoso, 1998).

3.4 Combining invariant feature subspaces and independent subspaces

Invariant-feature subspaces can be embedded in multidimensional independent component analysis by considering probability distributions for the n-tuples of s_i that are *spherically symmetric*, i.e. depend only on the norm. In other words, the probability density $p_j(.)$ of the n-tuple with index $j \in \{1, ..., J\}$, can be expressed as a function of the sum of the squares of the $s_i, i \in S_j$ only. For simplicity, we assume further that the $p_j(.)$ are equal for all j, i.e. for all subspaces.

This means that the logarithm of the likelihood L of the data, i.e. the K observed image patches $I_k(x, y), k = 1, ..., K$, given the model, can be expressed as

$$\begin{aligned} & \log L(I_k(x,y), k = 1...K; w_i(x,y), i = 1...m) \\ &= \sum_{k=1}^{K} \sum_{j=1}^{J} \log p(\sum_{i \in S_j} < w_i, I_k >^2) + K \log |\det \mathbf{W}| \end{aligned} \tag{5}$$

where $p(\sum_{i \in S_j} s_i^2) = p_j(s_i, i \in S_j)$ gives the probability density inside the j-th n-tuple of s_i.

Recall that prewhitening allows us to consider the $w_i(x,y)$ to be orthonormal, which implies that $\log|\det \mathbf{W}|$ is zero. This shows that the likelihood in Eq. (5) is a function of the norms of the projections of $I_k(x,y)$ on the subspaces indexed by j, which are spanned by the orthonormal basis sets given by $w_i(x,y), i \in S_j$. Since the norm of the projection of visual data on practically any subspace has a supergaussian distribution, we need to choose the probability density p in the model to be sparse (Olshausen and Field, 1996), i.e. supergaussian (Hyvärinen and Oja, 1997). For example, we could use the following probability distribution:

$$\log p(\sum_{i \in S_j} s_i^2) = -\alpha[\sum_{i \in S_j} s_i^2]^{1/2} + \beta, \tag{6}$$

which could be considered a multi-dimensional version of the exponential distibution (Field, 1994). The scaling constant α and the normalization constant β are determined so as to give a probability density that is compatible with the constraint of unit variance of the s_i, but they are irrelevant in the following. Thus we see that the estimation of the model consists of finding subspaces such that the *norms of the projections of the (whitened) data on those subspaces have maximally sparse distributions.*

The introduced 'independent feature subspace analysis' is a natural generalization of ordinary ICA. In fact, if the projections on the subspaces are reduced to dot-products, i.e. projections on 1-D subspaces, the model reduces to ordinary ICA, provided that, in addition, the independent components are assumed to have non-skewed distributions. It is to be expected that the norms of the projections on the subspaces represent some higher-order, invariant features. The exact nature of the invariances has not been specified in the model but will emerge from the input data, using only the prior information on their independence.

When independent feature subspace analysis is applied to natural image data, we can identify the norms of the projections $(\sum_{i \in S_j} s_i^2)^{1/2}$ as the responses of the complex cells. If the individual filter vectors $w_i(x,y)$ are identified with the receptive fields of simple cells, this can be interpreted as a hierarchical model where the complex cell response is computed from simple cell responses s_i, in a manner similar to the classical energy models for complex cells (Hubel and Wiesel, 1962; Pollen and Ronner, 1983; Heeger, 1992). It must be noted, however, that our model does not specify the particular basis of a given invariant-feature subspace.

3.5 Learning independent feature subspaces

Learning the independent feature subspace representation can be simply achieved by gradient ascent of the log-likelihood in Eq.(5). Due to whitening, we can constrain the vectors w_i to be orthogonal and of unit norm, as in ordinary ICA; these constraints usually speed up convergence. A stochastic gradient ascent of the log-likelihood can be obtained as

$$\Delta w_i(x,y) \propto I(x,y) < w_i, I > g(\sum_{r \in S_{j(i)}} < w_r, I >^2) \tag{7}$$

where $j(i)$ is the index of the subspace to which w_i belongs, and $g = p'/p$ is a nonlinear function that incorporates our information on the sparseness of the norms of the projections. For example, if we choose the distribution in Eq. (6), we have $g(u) = -\frac{1}{2}\alpha u^{-1/2}$, where the positive constant $\frac{1}{2}\alpha$ can be ignored. After every step of (7), the vectors w_i need to be orthonormalized; for a variety of methods to perform this, see (Hyvärinen and Oja, 1997; Karhunen et al., 1997).

The learning rule in (7) can be considered as 'modulated' nonlinear Hebbian learning. If the subspace which contains w_i were just one-dimensional, this learning rule would reduce to the learning rules for ordinary ICA given in (Hyvärinen and Oja, 1998) and closely related to those in (Bell and Sejnowski, 1997; Cardoso and Laheld, 1996; Karhunen et al., 1997). The difference is that in the general case, the Hebbian term is divided by a function of the output of the complex cell, given by $\sum_{r \in S_{j(i)}} < w_r, I >^2$, if we assume the terminology of the energy models. In other words, the Hebbian term is modulated by a top-down feedback signal. In addition to this modulation, the neurons interact in the form of the orthogonalizing feedback.

4 Experiments

To test our model, we used patches of natural images as input data $I_k(x, y)$, and estimated the model of independent feature subspace analysis.

4.1 Data and methods

The data was obtained by taking 16×16 pixel image patches at random locations from monochrome photographs depicting wild-life scenes (animals, meadows, forests, etc.). The images were taken directly from PhotoCDs, and are available on the World Wide Web[1]. The mean gray-scale value of each image patch (i.e. the DC component) was subtracted. The data was then low-pass filtered by reducing the dimension of the data vector by principal component analysis, retaining the 160 principal components with the largest variances. Next, the data was whitened by the zero-phase whitening filter, which means multiplying the data by $\mathbf{C}^{-1/2}$, where $\mathbf{C}$ is the covariance of the data (after PCA), see e.g. (Bell and Sejnowski, 1997). These preprocessing steps are essentially similar to those used in (Olshausen and Field, 1996; van Hateren and van der Schaaf, 1998). The likelihood in Eq. (5) for 50000 such observations was maximized under the constraint of orthonormality of the filters in the whitened space, using the averaged version of the learning rule in (7), i.e., we used the ordinary gradient of the likelihood instead of the stochastic gradient. The fact that the data was contained in a 160 dimensional subspace meant that the 160 basis vectors w_i now formed an orthonormal system for that subspace and not for the original space, but this did not necessitate any changes in the learning rule. The density p was chosen as in Eq. (6). The algorithm was initialized as in (Bell and Sejnowski, 1997) by taking as w_i the 160 middle columns of the identity matrix. We also tried random initial values for $\mathbf{W}$: these yielded qualitatively identical results, but using a localized filter set as the initial value improves considerably the convergence of the method, especially avoiding some of the filters getting stuck in local minima. This initialization lead, incidentally, to a weak topographical organization of the filters. The computations took about 10 hours on a single RISC processor. Experiments were made with different dimensions S_j for the subspaces: 2, 4, and 8 (in a single run, all the subspaces had the same dimension). The results shown below are for 4-dimensional subspaces, but the results are similar for other dimensions.

4.2 Results

Fig. 2 shows the filter sets of the 40 feature subspaces (complex cells), when subspace dimension was chosen to be 4. The results are shown in the zero-phase whitened space; note that due to orthogonality, the filters are equal to the basis vectors. The filters look qualitatively similar in the original, not whitened space. The only difference is that in the original space, the filters are sharper, i.e. more concentrated on higher frequencies.

[1] WWW address: `http://www.cis.hut.fi/~phoyer/NCimages.html`

It can be seen that the linear filters associated with a single complex cell all have approximately the same orientation and frequency. Their locations are not identical, but close to each other. The phases differ considerably. Every feature subspace can thus be considered a generalization of a quadrature-phase filter pair as found in the classical energy models (Pollen and Ronner, 1983), enabling the cell to be selective to given orientation and frequency, but invariant to phase and somewhat invariant to shifts. Using 4 filters instead of a pair greatly enhances the shift invariance of the feature subspace. In fact, when the subspace dimension was 2, we obtained approximately quadrature-phase filter pairs.

To quantitatively demonstrate the properties of the model we compared the responses of a representative feature subspace and the associated linear filters, for different stimulus configurations. First, an optimal stimulus for the feature subspace was computed in the set of Gabor filters. One of the stimulus parameters was changed at a time to see how the response changes, while the other parameters were held constant at the optimal values. Some typical simuli are depicted in Fig. 3. The investigated parameters were phase, orientation, and location (shift).

Fig. 4 shows the results for one typical feature subspace. The 4 linear filters spanning the feature subspace are shown in Fig. 4 a). The optimal stimulus values (for the feature subspace) are represented by 0 in these results, i.e. the values given here are departures from the optimal values. The responses are in arbitrary units. For different phases, ranging from $-\pi/2$ to $\pi/2$ we thus obtained Fig. 4 b). On the bottom row, we have the response curve of the feature subspace; for comparison, the absolute values of the responses of the associated linear filters are also shown in the 4 plots above it. Similarly, we obtained c) for different locations (location was moved along the direction perpendicular to the preferred orientation; the shift units are arbitrary), and d) for different orientations (ranging from $-\pi/2$ to $\pi/2$). The response curve of the feature subspace shows clear invariance with respect to phase and location; note that the response curve for location consists of an approximately gaussian envelope, as observed in complex cells (von der Heydt, 1995). In contrast, no invariance for orientation was observed. The responses of the linear filters, on the other hand, show no invariances with respect to any of these parameters, which is in fact a necessary consequence of the linearity of the filters (Pollen and Ronner, 1983; von der Heydt, 1995). Thus this feature subspace shows the desired properties of phase invariance and limited shift invariance, contrasting them with the properties of the underlying linear filters.

To see how the above results hold for the whole population of feature subspaces, we computed the same response curves for all the feature subspaces, and for comparison, for all the underlying linear filters. Optimal stimuli were separately computed for all the feature subspaces. In contrast to the above results, we computed the optimal stimuli separately for each linear filter as well: this facilitates the quantitative comparison. The response values were normalized so that the maximum response for each subspace or linear filter was equal to 1. Figure 5 shows the responses for 10 typical feature subspaces and 10 typical linear filters, whereas Fig. 6 shows the median responses of the whole populations of 40 feature subspaces and 160 linear filters, together with the 10% and 90% percentiles.

In Fig. 5 a) and Fig. 6 a) , the responses are given for varying phase. The top row shows the absolute responses of the linear filters, and in the bottom row the corresponding results for the feature subspaces are depicted. The figures show that phase invariance is a strong property of the feature subspaces: the minimum response was usually at least two thirds of the maximum response. Fig. 5 b) and Fig. 6 b) show the results for location shift. Clearly, the 'receptive field' of a typical feature subspace is larger and more invariant than that of a typical linear filter. As for orientation, Fig. 5 c) and Fig. 6 c) depict the corresponding results, showing that the orientation selectivity was approximately equivalent in linear filters and feature subspaces. Thus we see that invariances with respect to translation and especially phase, as well as orientation selectivity, hold for the population of feature subspaces in general.

5 Discussion

An approach closely related to ours is given by the adaptive subspace self-organizing map (Kohonen, 1996), which is based on competitive learning of a invariant-feature subspace representation similar to ours. Using two-dimensional subspaces, an emergence of filter pairs with phase invariance and limited shift invariance was shown in (Kohonen, 1996). However, the emergence of shift invariance in (Kohonen, 1996) was conditional to restricting consecutive patches to come from nearby locations in the image, giving the input data a temporal structure like in a smoothly changing image sequence. Similar developments were given by Földiák (1991). In contrast to these two theories, we formulated an explicit image model, showing that emergence is possible using patches at random, independently selected locations, which proves that there is enough information in static images to explain the properties of complex cells.

In conclusion, we emphasize that we have provided a model of the emergence of phase and shift invariant features, i.e. the principal properties of visual complex cells, using the very same principle of maximization of independence (Comon, 1994; Barlow, 1994; Field, 1994) that has been used to explain simple cell properties (Olshausen and Field, 1996). Maximizing the independence, or equivalently the sparsenesses, of linear filter outputs, the model gives simple cell properties. Maximizing the independence of the *norms of the projections* on linear subspaces, complex cell properties emerge. This provides further evidence for the fundamental importance of dependence reduction as a strategy for sensory information processing. It is, in fact, closely related to such fundamental objectives as minimizing code length, and reducing redundancy (Barlow, 1989; Barlow, 1994; Field, 1994; Olshausen and Field, 1996). It remains to be seen if our framework could also account for the movement-related, binocular, and chromatic properties of complex cells; the simple cell model has had some success in this respect (van Hateren and Ruderman, 1998). Extending the model to include overcomplete bases (Olshausen and Field, 1997) may be useful for such purposes.

References

Barlow, H. (1989). Unsupervised learning. *Neural Computation*, 1:295–311.

Barlow, H. (1994). What is the computational goal of the neocortex ? In Koch, C. and Davis, J., editors, *Large-scale neuronal theories of the brain*. MIT Press, Cambridge, MA.

Bell, A. and Sejnowski, T. (1997). The 'independent components' of natural scenes are edge filters. *Vision Research*, 37:3327–3338.

Cardoso, J. F. (1998). Multidimensional independent component analysis. In *Proc. ICASSP'98*, Seattle, WA.

Cardoso, J.-F. and Laheld, B. H. (1996). Equivariant adaptive source separation. *IEEE Trans. on Signal Processing*, 44(12):3017–3030.

Comon, P. (1994). Independent component analysis – a new concept? *Signal Processing*, 36:287–314.

Field, D. (1994). What is the goal of sensory coding? *Neural Computation*, 6:559–601.

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, 3:194–200.

Heeger, D. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–198.

Hubel, D. and Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 73:218–226.

Hyvärinen, A. and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492.

Hyvärinen, A. and Oja, E. (1998). Independent component analysis by general nonlinear Hebbian-like learning rules. *Signal Processing*, 64(3):301–313.

Karhunen, J., Oja, E., Wang, L., Vigario, R., and Joutsensalo, J. (1997). A class of neural networks for independent component analysis. *IEEE Trans. on Neural Networks*, 8(3):486–504.

Kohonen, T. (1995). *Self-Organizing Maps.* Springer-Verlag, Berlin, Heidelberg, New York.

Kohonen, T. (1996). Emergence of invariant-feature detectors in the adaptive-subspace self-organizing map. *Biological Cybernetics*, 75:281–291.

Mel, B. W., Ruderman, D. L., and Archie, K. A. (1998). Translation-invariant orientation tuning in visual 'complex' cells could derive from intradendritic computations. *Journal of Neuroscience*, 18:4325–4334.

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609.

Olshausen, B. A. and Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37:3311–3325.

Pham, D.-T., Garrat, P., and Jutten, C. (1992). Separation of a mixture of independent sources through a maximum likelihood approach. In *Proc. EUSIPCO*, pages 771–774.

Pollen, D. and Ronner, S. (1983). Visual cortical neurons as localized spatial frequency filters. *IEEE Trans. on Systems, Man, and Cybernetics*, 13:907–916.

van Hateren, J. H. and Ruderman, D. L. (1998). Independent component analysis of natural image sequences yields spatiotemporal filters similar to simple cells in primary visual cortex. *Proc. Royal Society ser. B*, 265:2315–2320.

van Hateren, J. H. and van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Royal Society ser. B*, 265:359–366.

von der Heydt, R. (1995). Form analysis in visual cortex. In Gazzaniga, M. S., editor, *The Cognitive Neurosciences*, pages 365–382. MIT Press.

von der Malsburg, C., Shams, L., and Eysel, U. (1998). Recognition of images from complex cell responses. Abstract at Proc. of Society for Neuroscience Meeting.

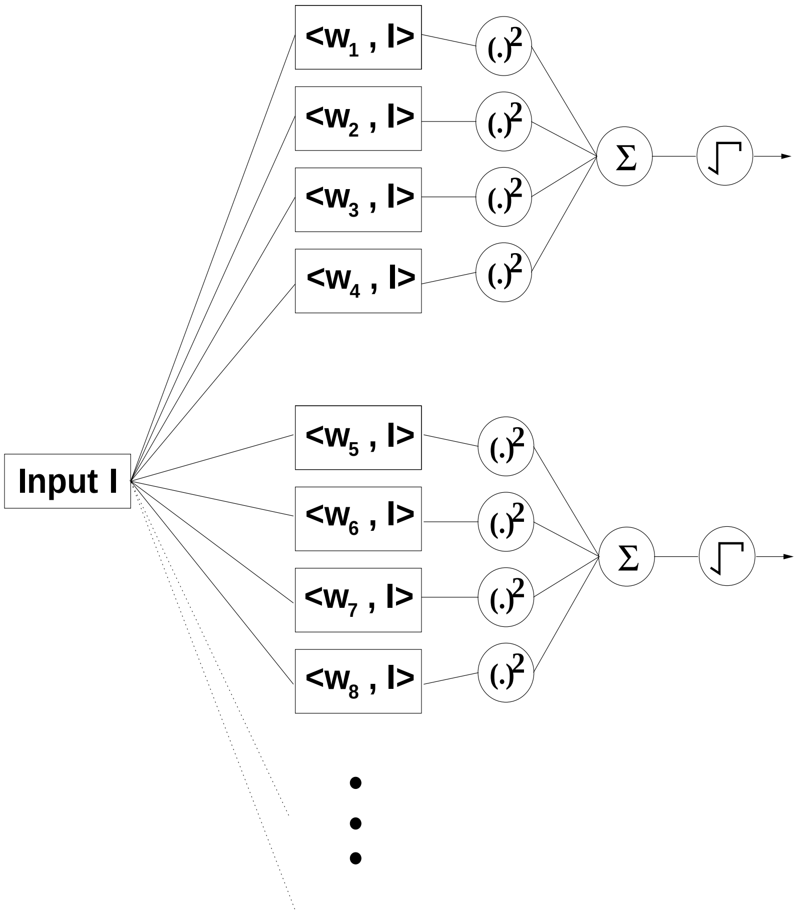

Figure 1: A graphical depiction of the feature subspaces. First, dot products of the input data with a set of basis vectors are taken. Here, we have two subspaces with 4 basis vectors in each. The dot products are then squared, and the sums are taken inside the feature subspaces. Thus we obtain the (squares of the) norms of the projections on the feature subspaces, which are considered as the responses of the subspaces. Square roots may be taken for normalization. This scheme represents features that go beyond simple linear filters, possibly obtaining invariance with respect to some transformations of the input, for example shift and phase invariance. The subspaces, or the underlying vectors $\mathbf{w}_i$, may be learned by the principle of maximum sparseness, which coincides here with maximum likelihood of a generative model.

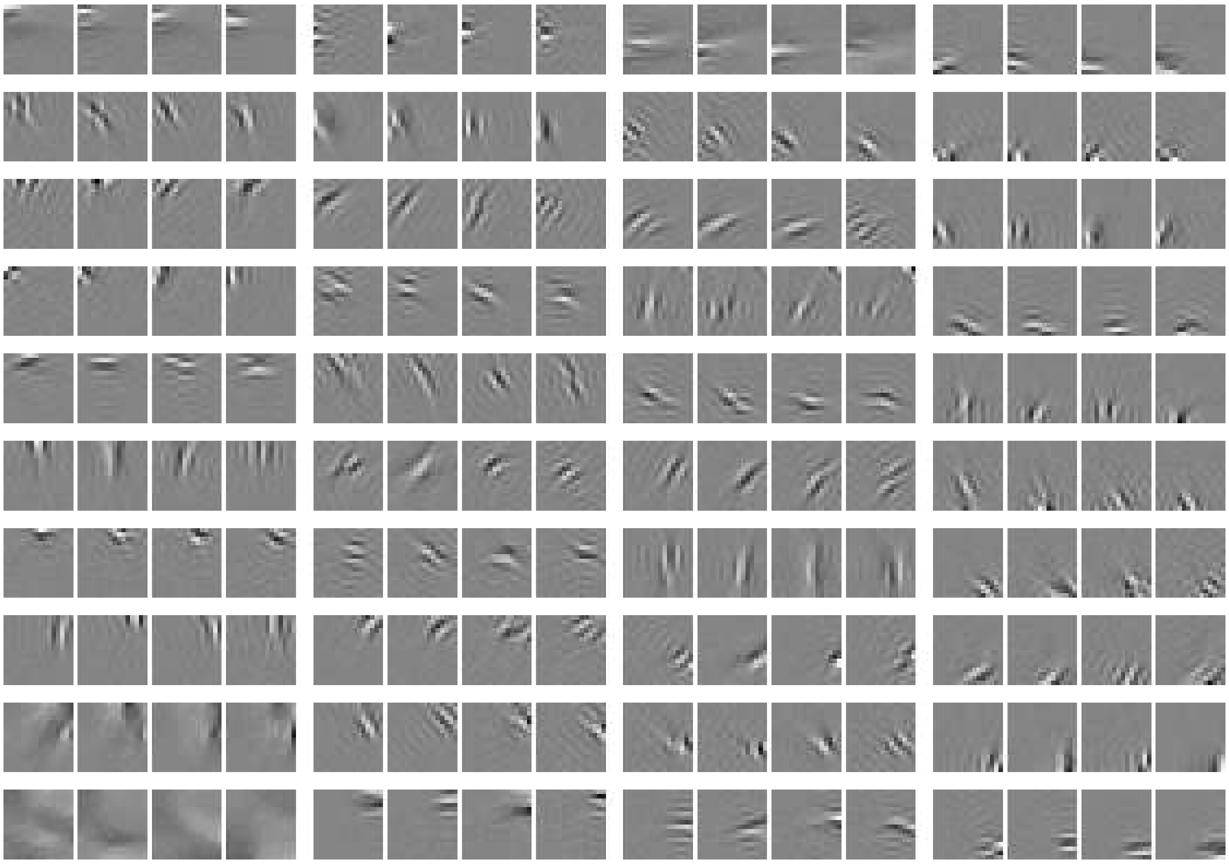

Figure 2: The linear filter sets associated with the feature subspaces (model complex cells), as estimated from natural image data. Every group of four filters spans a single feature subspace (for whitened data).

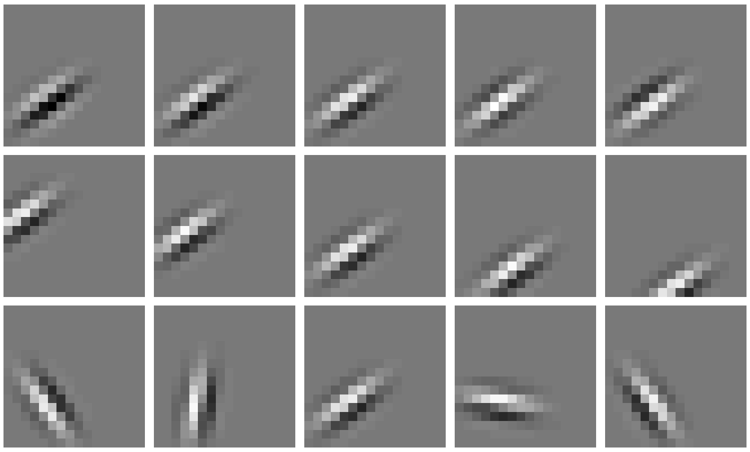

Figure 3: Typical stimuli used in the experiments in Figs. 4 to Fig. 6 below. The middle column shows the optimal Gabor stimulus. One of the parameters was varied at a time. Top row: varying phase. Middle row: varying location (shift). Bottom row: varying orientation.

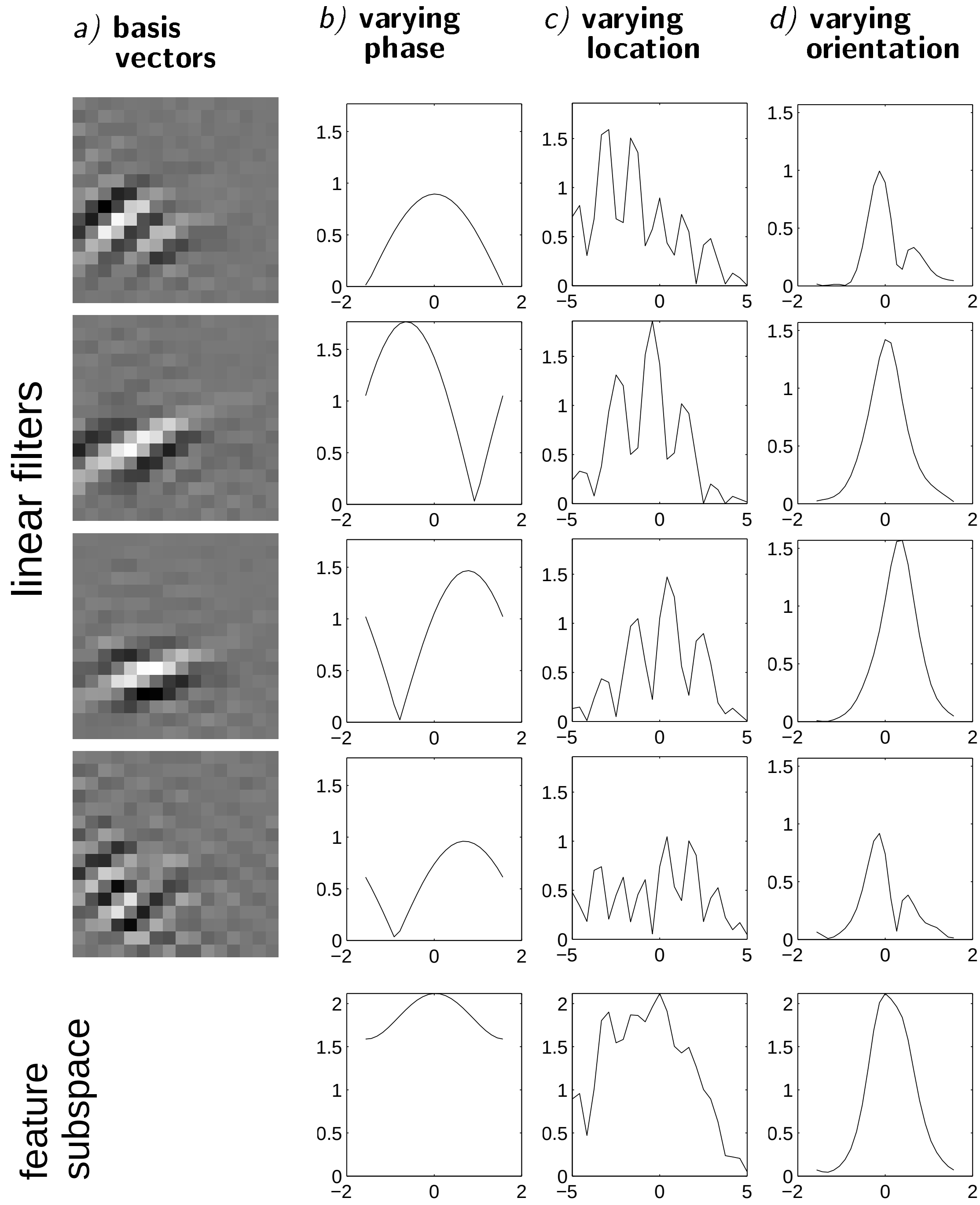

Figure 4: Responses elicited from a feature subspace and the underlying linear filters by different stimuli, for stimuli as in Fig. 3. a) the 4 (arbitrarily chosen) filters spanning the feature subspace. b) Effect of varying phase. Upper 4 rows: absolute responses of linear filters (simple cells) to stimuli of different phases. Bottom row: response of feature subspace (complex cell). c) Effect of varying location (shift), as in b). d) Effect of varying orientation, as in b).

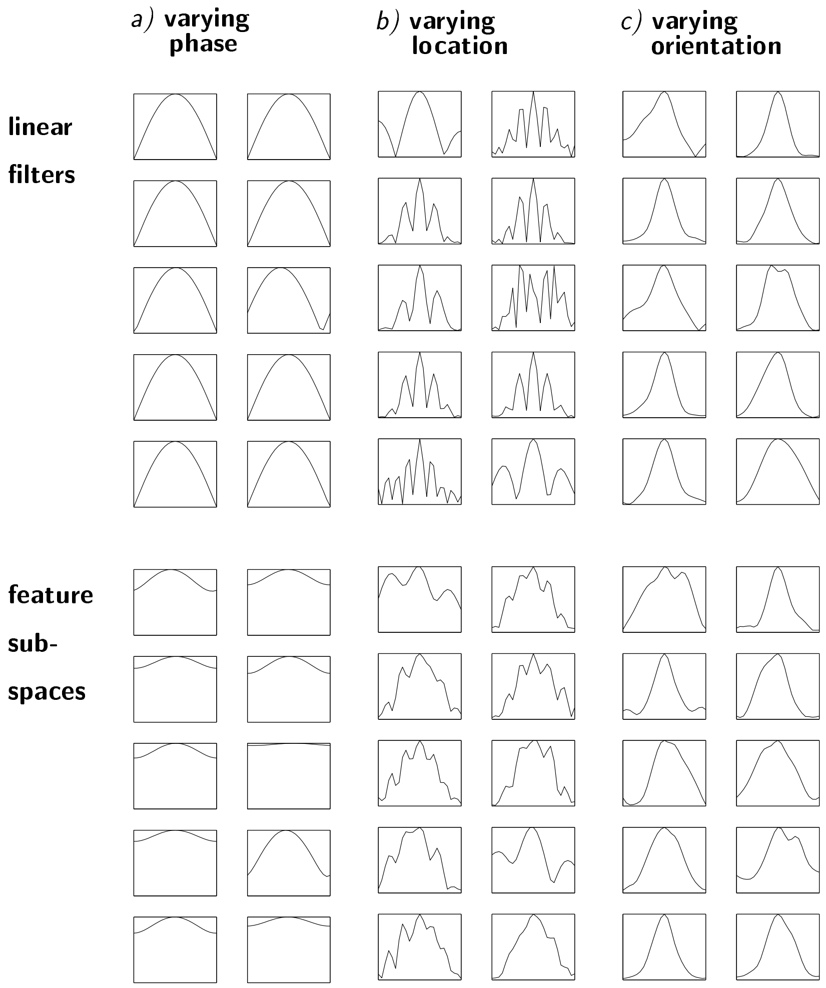

Figure 5: Some typical response curves of the feature subspaces and the underlying linear filters when one parameter of the optimal stimulus is varied, as in Fig. 3. Top row: responses (in absolute values) of 10 linear filters (simple cells). Bottom row: responses of 10 feature subspaces (complex cells). a) Effect of varying phase. b) Effect of varying location (shift). c) Effect of varying orientation.

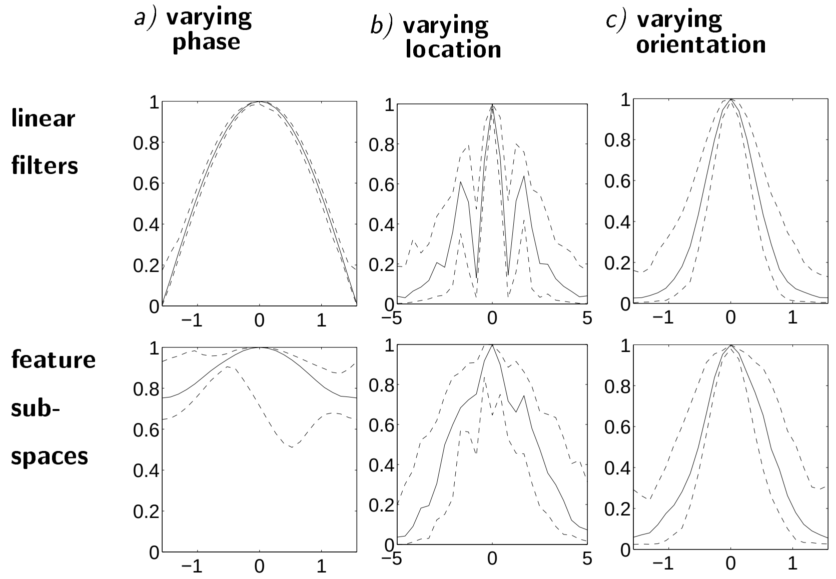

Figure 6: Statistical analysis of the properties of the whole population of feature subspaces, with the corresponding results for linear filters given for comparison. In all plots, the solid line gives the median response in the population of all cells (filters or subspaces), and the dotted lines give the 90% and 10% percentiles of the responses. Stimuli were as in Fig. 3. Top row: responses (in absolute values) of linear filters (simple cells). Bottom row: responses of feature subspaces (complex cells). a) Effect of varying phase. b) Effect of varying location (shift). c) Effect of varying orientation.